NONPARAMETRIC MARKOV PRIORS FOR TISSUE SEGMENTATION

Zhuang Song, Suyash P. Awate, James C. Gee

Penn Image Computing and Science Lab
University of Pennsylvania, Philadelphia, PA 19104, U.S.A

ABSTRACT

This paper presents a novel method to construct a probabilistic tissue prior, for Bayesian tissue segmentation, which is based on *nonparametric Markov statistics* of tissue intensities learned from training data. The proposed nonparametric Markov (NPM) prior is in contrast to the conventional tissue-probability-map (TPM) prior that is based on the voxel location in a common anatomical template space. Given a set of manually labeled voxels as the training set, the NPM prior is constructed by learning a fuzzy classification function that distinguishes the Markov statistics of tissue intensities in a statistical supervised-learning framework. The validation experiments in this paper compare the efficacy of the NPM prior to that of the TPM prior in producing tissue segmentations, and demonstrate the advantages of the NPM prior, qualitatively and quantitatively, over the TPM prior, especially in cortical regions.

Index Terms— image segmentation, probability, learning systems, magnetic resonance imaging

1. INTRODUCTION

Tissue segmentation remains a challenging task in biomedical image analysis due to complex biological tissue properties and limitations of imaging technologies. Bayesian segmentation approaches have been widely applied to biomedical images. Their popularity stems from the rigorous incorporation of priors, gained from clinical and biological knowledge and understanding of the MR imaging technique, to tune tissue probabilities and constrain the segmentation solution. A widely used prior is the template-based *tissue probability map* (TPM), or probabilistic atlas, which gives the tissue probabilities based on the anatomical location in the image [1]. Because of the significant variability in anatomical structures, especially in the cortical regions where there may *not* be a one-to-one correspondence between two individuals, the quality of the image registration may be compromised. Furthermore, the template is typically produced by solving an optimization problem in a very high-dimensional space (the anatomical image) [2] with a small number of data points (the manually segmented image) relative to the dimension of the space. As a result, there may be significant blurring in the TPM. Apart from being time consuming, reliable manual segmentation of certain datasets, is challenging to obtain. Labeling errors in the training data will propagate to the constructed TPM prior. These factors, taken together, undermine the efficacy of TPM priors for tissue segmentation.

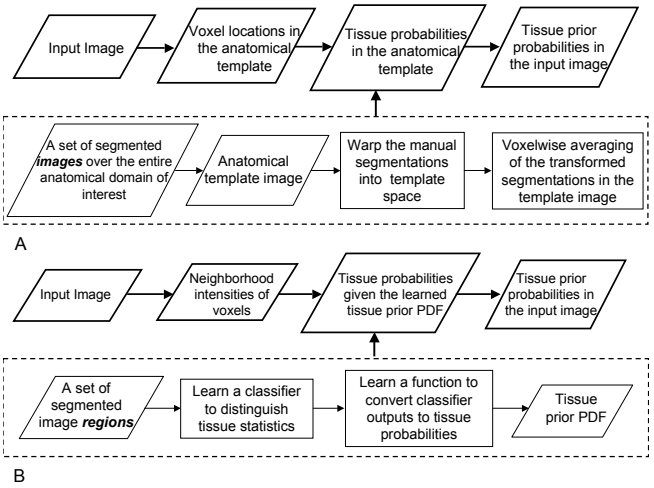

Fig. 1. Construction process of (A) the conventional TPM prior and (B) the NPM prior. The dashed frames indicate the prior learning processes.

We propose an intensity-based nonparametric Markov (NPM) prior for tissue segmentation. The prior is learned from nonparametric high-order statistics, using a supervised learning framework, *i.e.* support vector machine (SVM) [3], although the proposed method can be adapted to use any other supervised-learning framework that performs function approximation. Figure 1 compares the construction of the NPM prior with the TPM prior. One prominent advantage of the NPM prior is that it does *not* require a large number of manual segmentations of the entire anatomy represented in the image volumes, as required by the TPM prior in order to capture the tissue-structure variability. In principle, it is much easier to capture the variability in the space defined on neighborhood intensities than to capture the variability over the entire anatomical image space, because the latter space has many more dimensions.

The validation experiments demonstrate that the proposed NPM prior can provide estimates of tissue probabilities with significantly greater discriminatory power in the highly-folded cortex, in contrast to that of the TPM prior. Because of the error tolerance of the fuzzy learning method, the manual segmentation need not be of high quality. The supervised learning framework underlying the NPM prior makes it robust against image noise and mislabeling in manual segmentation in the training images.

2. RELATED WORK

The NPM prior models context dependence using estimates of joint intensity PDFs in voxel neighborhoods. Considering the neighborhood intensities as a realization of a Markov random field (MRF), the inherent structure of their distribution is equivalent to the Markov probability density function (PDF). This approach is closely related to the research on learning the Markov statistics of natural images [4], and empirical MRF modeling for texture synthesis [5] and denoising [6]. Awate et al. [7] employ kernel density estimation for unsupervised learning of the Markov PDFs in images for segmenting tissues in adult-brain MR images. Different than these nonparametric modeling methods that explicitly learn the Markov statistics of classes (generative learning) [6, 7], the proposed NPM prior is constructed by directly learning a classifier to distinguish the classes (discriminative learning) [3].

3. BAYESIAN IMAGE SEGMENTATION AND PRIOR PROBABILITIES

Image segmentation can be formulated as a Bayesian inference problem in a random-field framework. Given a set of voxels $\mathbf{T}$ on a Cartesian grid, a random field image model, *i.e.* a set of observed random variables $\mathbb{X} = \{X_t\}_{t\in\mathbf{T}}$, is associated with a set of hidden random variables $\mathbb{L} = \{L_t\}_{t\in\mathbf{T}}$ where L_t is assigned as tissue types in the segmentation task. An input image $\mathbf{X}$ consisting of a set of single or multi-valued data $\{x_t\}_{t\in\mathbf{T}}$, which is a realization of the random field $\mathbb{X}$, is associated with a particular segmentation $\mathbf{L} = \{l_t\}_{t\in\mathbf{T}}$. To introduce the intensity-based tissue prior, we draw a concept of the synthesized image from the texture analysis literature. A novel image can be synthesized by a learned statistical model. Joint shape and appearance prior models have been applied to atlas matching [8] and segmentation [9] of medical images, by minimizing the difference between the target image and the synthesized model image. Similar to the formulation proposed in [9], a MAP segmentation can be given by

$$\begin{aligned}\hat{\mathbf{L}}, \hat{\mathbf{X}}_{\mathbf{L}} &= \underset{\mathbf{L},\mathbf{X}_{\mathbf{L}}}{\operatorname{argmax}}\ P(\mathbf{L}, \mathbf{X}_{\mathbf{L}}|\mathbf{X}) \\ &= \underset{\mathbf{L},\mathbf{X}_{\mathbf{L}}}{\operatorname{argmax}}\ P(\mathbf{X}|\mathbf{L}, \mathbf{X}_{\mathbf{L}})\tilde{P}(\mathbf{L}, \mathbf{X}_{\mathbf{L}}),\end{aligned}$$

where $\mathbf{X}_{\mathbf{L}}$ is the synthetic model image of the real image $\mathbf{X}$, $\hat{\mathbf{X}}_{\mathbf{L}}$ the optimal estimation of the synthetic image $\mathbf{X}_{\mathbf{L}}$, and $\hat{\mathbf{L}}$ the optimal segmentation. To avoid confusion, we use $\tilde{P}$ to denote the prior probability. Because we are only interested in segmentation, it is reasonable to assume that the synthetic image is very close to the real image, *i.e.* $\mathbf{X}_{\mathbf{L}} \approx \mathbf{X}$, therefore

$$\hat{\mathbf{L}} \approx \underset{\mathbf{L}}{\operatorname{argmax}}\ P(\mathbf{X}|\mathbf{L}, \mathbf{X}_{\mathbf{L}} \approx \mathbf{X})\tilde{P}(\mathbf{L}, \mathbf{X}), \qquad (1)$$

where $P(\mathbf{X}|\mathbf{L}, \mathbf{X}_{\mathbf{L}} \approx \mathbf{X}) \approx P(\mathbf{X}|\mathbf{L})$ is the likelihood function, and $\tilde{P}(\mathbf{L}, \mathbf{X})$ the joint label and intensity prior density function.

4. LEARNING NPM PRIORS VIA SUPPORT VECTOR MACHINES

The joint label and intensity prior density function $\tilde{P}(\mathbf{L}, \mathbf{X})$ in (1) can be regarded as a product of the different priors, for instance template-based tissue priors or MRF smoothness priors. In this paper we are interested in the estimation of the NPM prior relying on high-order intensity statistics in the training data, which we denote as $\tilde{P}_X(\mathbf{X}, \mathbf{L})$. To learn the proposed prior we use the supervised learning framework of the SVM [3]. To generate tissue prior probabilities for a novel image, the SVM model is applied to the novel image and the decision value at each voxel is subsequently converted to estimate the prior probabilities. The proposed prior relies on high-order intensity statistics as the distinguishing feature. The choice of the neighborhood is user-dependent. The results in this paper employ a first-order neighborhood in 3D images. The proposed approach can be extended in a straightforward way to incorporate other features to discriminate between tissues.

Assume the conditional independence between data and use the neighborhood intensities of each image voxel as the feature vector,

$$\tilde{P}_X(\mathbf{X}, \mathbf{L}) = \prod_{t\in\mathbf{T}} \tilde{P}_X(\mathbf{y}_t, l_t),$$

where $\mathbf{y}_t = \{x_i, i \in \mathbf{N}_t\}$, $\mathbf{N}_t$ is the neighborhood of voxel t. For an input datum $\mathbf{y}_t$ whose label l_t is to be determined, the SVM produces a decision value $f_{SVM}(\mathbf{y}_t)$ that is equal to the signed distance of the point $\mathbf{y}_t$ from the decision hypersurface. The literature presents several ways of calibrating the decision value $f_{SVM}(\mathbf{y}_t)$ to a probability value that can be employed in a Bayesian classification scheme. We adopt a popular technique proposed by Platt [10], which first fits a sigmoid function

$$\tilde{P}_X(\mathbf{y}_t, l_t) = \frac{1}{1 + \exp(Af_{SVM}(\mathbf{y}_t) + B)}$$

to the class probabilities, obtained for the SVM outputs. The optimal values for the parameters A and B are learned by a

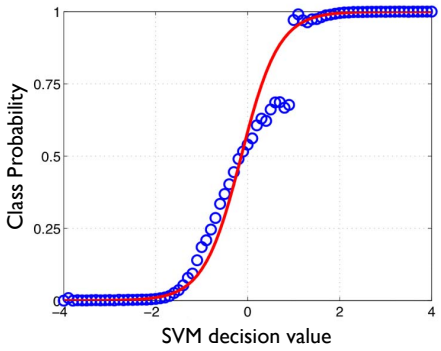

Fig. 2. SVM calibration: fitting a sigmoid function for the GM tissue class in simulated BrainWeb MR images. Each circle denotes the class probability computed from the particular SVM decision value [10].

maximum-likelihood method from the training set [10]. In order to avoid bias, the training set used to learn the calibration function should be different from that used to learn the SVM classifier. Figure 2 depicts the fitting of the calibration function to the probability of tissue classification of a simulated brain MR dataset. The probability estimation via SVM can solve multi-class problems [11].

The key idea underlying SVMs is to project the data into a high-dimensional feature space via a kernel function, where the projected data is assumed to be separable using a linear classifier (a decision hyperplane). Subsequently, SVMs find an optimal linear classifier in the projected high-dimensional space that maximizes the separation between the classes. The optimization finding a linear classifier in the projected high-dimensional space is a quadratic-programming problem involving inner products of feature vectors. Thus, fast algorithms can be brought to bear on its solution.

We choose a radial basis function (RBF) as the kernel function that implicitly projects the data into an infinite-dimensional Hilbert space. The infinite-dimensional space helps to effectively handle the complexity in the medical image data. Sensitivity of nonlinear SVMs to outliers, or noise, can be reduced by allowing some training data to be misclassified. Such conditions are typically incorporated in the optimization framework through slack variables associated with each training data. Fuzzy SVMs reduce overfitting of the classifier and make more robust the training process to (i) image noise and (ii) manual segmentation errors in the training data set.

5. EVALUATION

We demonstrate the efficacy of the NPM prior by validating the segmentations produced by *using the prior term alone*, and making the likelihood term noninformative. Specifically, we define the optimal segmentation at each voxel t as follows:

$$l_t^* = \underset{l_t}{\operatorname{argmax}}\ \tilde{P}(\mathbf{y}_t|l_t). \tag{2}$$

We compared the results of the NPM prior with those of the TPM prior. We constructed TPM priors using a state-of-the-art diffeomorphic nonlinear registration method [2]. For quantitative comparisons of the segmentations, we used the Dice overlap metric.

The efficacy of the proposed NPM prior depends on the consistency of MR image appearance of the tissues. However, MR intensities in multiple scans of a body region of a single patient using the same scanner and the imaging protocol do *not* guarantee the same intensity value. Intensity artifacts, noise, and partial-volume effects may reduce this consistency, in addition to developmental and pathological conditions. To address this problem, we ensure a reasonable match of the histograms of the images in the training set.

We present results on adult-brain MR images from the IBSR repository in order to evaluate the performance of the NPM prior in the presence of imaging artifacts and brain-structure variability actually encountered in practice. The IBSR repository comprises 17 T1-weighted adult brain MR images along with the manual tissue segmentations. The image voxel size was $1 \times 1 \times 1.5$ mm^3. Note that inhomogeneity correction was not performed for the data. We randomly selected 8 images as the training set to construct the TPM prior and the NPM prior, and the rest as test images.

It is evident in Figure 3 that the NPM prior was able to accurately discriminate between tissues in the cortex, where the TPM prior provided weak estimates of tissue probabilities. On the other hand, in some other regions where the inter-tissue intensity contrast was extremely low, e.g. in the deep GM, the location-based TPM prior was able to perform better as they learned these locations from manual segmentation. Figure 4 shows the quantitative comparison using the Dice overlaps of the segmentation with ground truth segmentation. For GM, the average ($\pm$ SD) Dice overlaps with the ground truth were $0.891(\pm 0.025)$ for the TPM-based segmentation and $0.925(\pm 0.020)$ for the NPM-based segmentation ($p < 0.02$ for the paired t-test). For WM, the average ($\pm$ SD) Dice overlaps with the ground truth were $0.836(\pm 0.015)$ for the TPM-based segmentation and $0.865(\pm 0.045)$ for the NPM-based segmentation ($p < 0.08$ for the paired t-test). The difference between the performance of the two priors can be more clearly seen by considering the average Dice overlap for GM and WM together. In this case, the average Dice overlaps with the ground truth segmentation were 0.863 ± 0.034 for the TPM-based segmentation and 0.895 ± 0.046 for the NPM-based segmentation ($p < 0.02$ for the paired t-test).

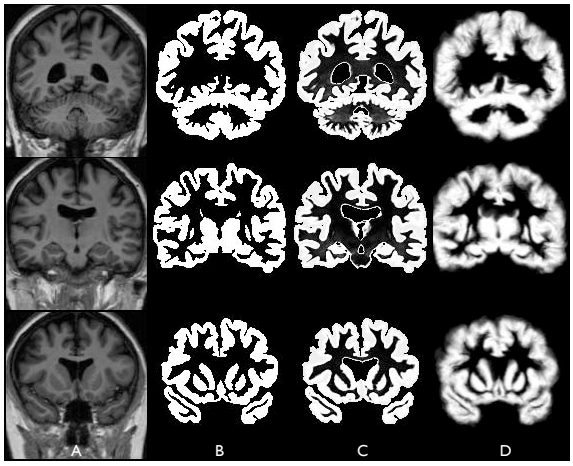

Fig. 3. (A) Coronal slices of one IBSR image. (B) The ground truth for the GM segmentation. The probabilities for GM derived from (C) the NPM prior and (D) the TPM prior.

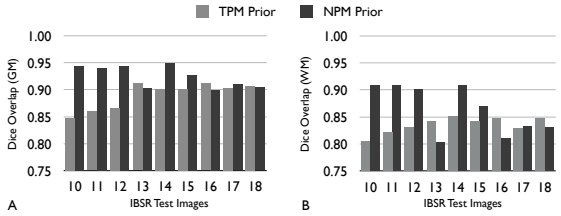

Fig. 4. Quantitative comparison between the intensity-based NPM prior and the location-based TPM prior through segmentations of test images from the IBSR repository. (A) and (B) show the Dice overlap with the ground truth for GM and WM, respectively.

6. DISCUSSION AND CONCLUSIONS

We propose an intensity-based NPM prior, which distinguishes between the Markov statistics of tissue intensities. The NPM prior exploits the discriminative power of fuzzy nonlinear SVMs to learn the complex class boundaries in the feature space comprising of neighborhood intensities. The evaluation experiments on real brain-MR images show that the NPM prior has significantly greater discriminatory power than the template-based TPM prior in the cortex. On the other hand, the location-based TPM prior can provide valuable information in regions where the imaging produces too *low* contrast between tissue types, e.g. subcortical regions, where the TPM prior can effectively complement the intensity-based NPM prior. Future work will deal with combining the strengths of the two types of priors to facilitate reliable tissue segmentation.

The segmentations in this paper have been produced using the *prior terms alone*. A complete segmentation framework will include a likelihood term as well that models the statistics of tissue intensities in the image to be segmented. On the other hand, by providing a better estimate of the initial segmentation, the NPM prior can lead to faster convergence.

7. REFERENCES

[1] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, pp. 839–51, 2005.

[2] B. B. Avants and J. C. Gee, "Geodesic estimation for large deformation anatomical shape and intensity averaging," *Neuroimage*, vol. Suppl. 1, pp. S139–150, 2004.

[3] V. N. Vapnik, *Statistical Learning Theory*, Wiley: New York, 1998.

[4] S. C. Zhu and D. Mumford, "Prior learning and Gibbs reaction-diffusion," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 19, no. 11, pp. 1236–50, 1997.

[5] J. Portilla and E. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vision*, vol. 40, pp. 49–71, 2000.

[6] K. Popat and R. Picard, "Cluster based probability model and its application to image and texture processing," *IEEE Trans. Image Proc.*, vol. 6, no. 2, pp. 268–84, 1997.

[7] S. P. Awate, T. Tasdizen, R. L. Foster, and R. T. Whitaker, "Adaptive Markov modeling for mutual-information-based unsupervised MRI brain-tissue classification," *Medical Image Analysis*, vol. 10, no. 5, pp. 726–39, 2006.

[8] T. F. Cootes, C. Beeston, G. J. Edwards, and C. J. Taylor, "A unified framework for atlas matching using active appearance models," in *Proceedings of the 16th International Conference on Information Processing in Medical Imaging (IPMI)*, London, UK, 1999, pp. 322–33, Springer-Verlag.

[9] J. Yang and J. S. Duncan, "3D image segmentation of deformable objects with joint shape-intensity prior models using level sets," *Med. Image Anal.*, vol. 8, no. 3, pp. 285–94, 2004.

[10] J. C. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," in *Advances in Large Margin Classifiers*, A.J. Smola, P.L. Bartlett, B. Scholkopf, and D. Schuurmans, Eds. MIT Press, 2000.

[11] T. Wu, C. Lin, and R. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.